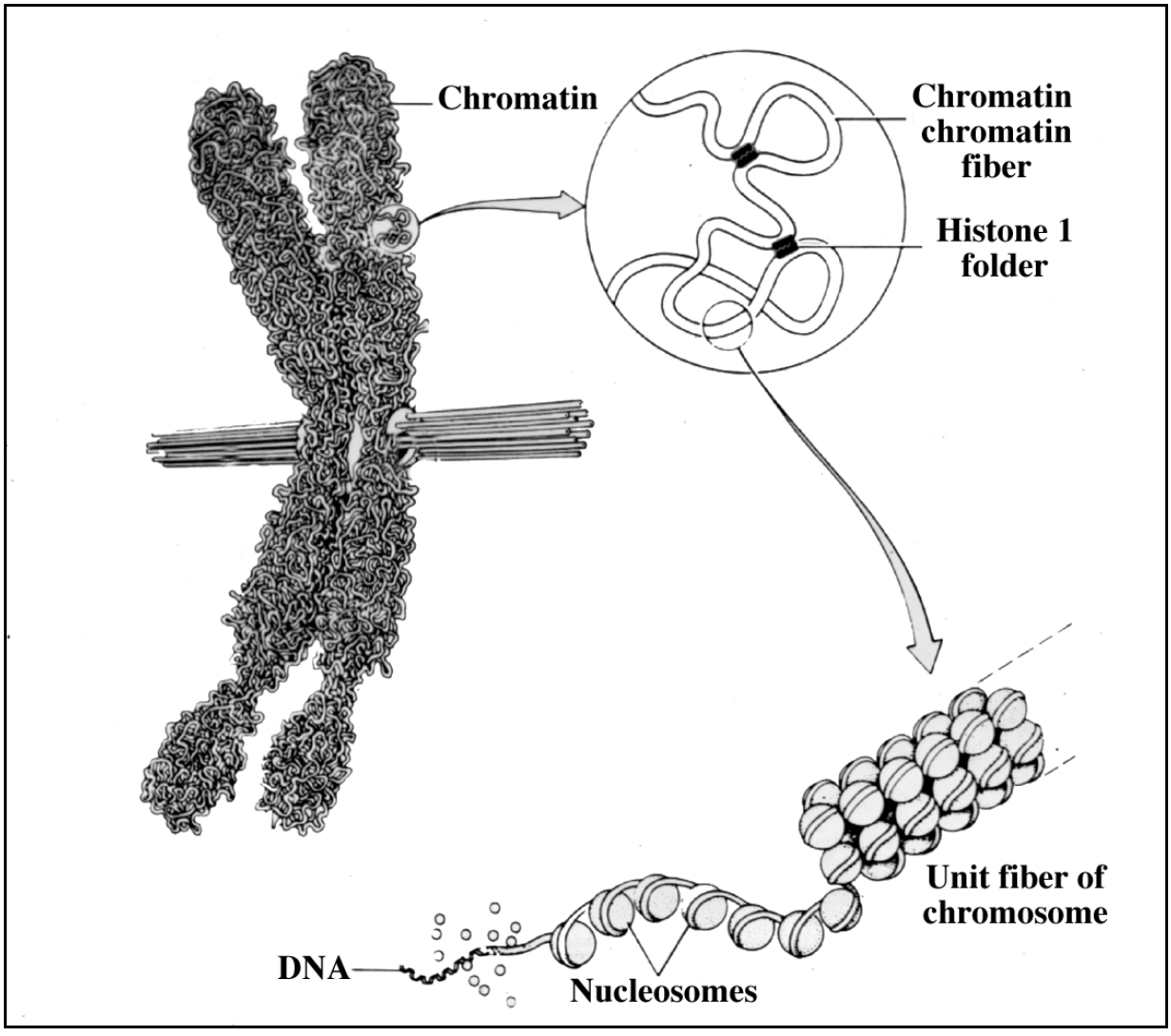
GENE STRUCTURE

Eukaryotic genes consist of sequences of DNA which are genuine coding sequences (exons) and intervening non-coding sequences (introns). In addition there are additional non-coding sequences of DNA between genes. It is estimated that only about 5% of the human genome contains actual coding sequences. Hundreds to thousands of copies of repetitive sequences occur both within genes, possibly representing multiple copies of highly used genes, and between genes, possibly serving a spacers. Their exact function is unknown. One commonly occurring highly repetitive sequence is the Alu sequence (Alu element) consisting of 300,000 to 500,000 base pairs scattered throughout the human genome.

DNA is tightly coiled (supercoiled) and packed in chromosomes. Chromatin refers to DNA sequences coiled around histone proteins.



Double stranded DNA sequences approximately 200 base pairs in length are wrapped around 9 histones forming a nucleosome. Chromatin actually consists of multiple nucleosomes.

The small single histone is H1 while the 8 clustered histones represent two copies each of H2A, H2B, H3 and H4.

During DNA replication, histones become accessible for in situ histone staining & this feature may be used as an indication of active cell proliferation.

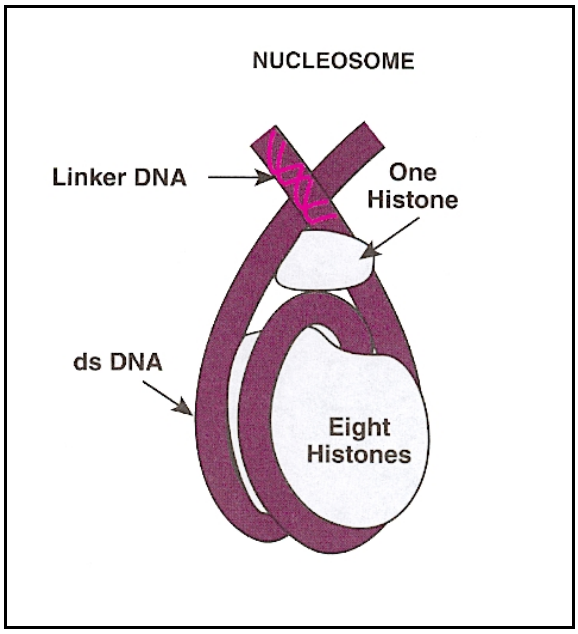
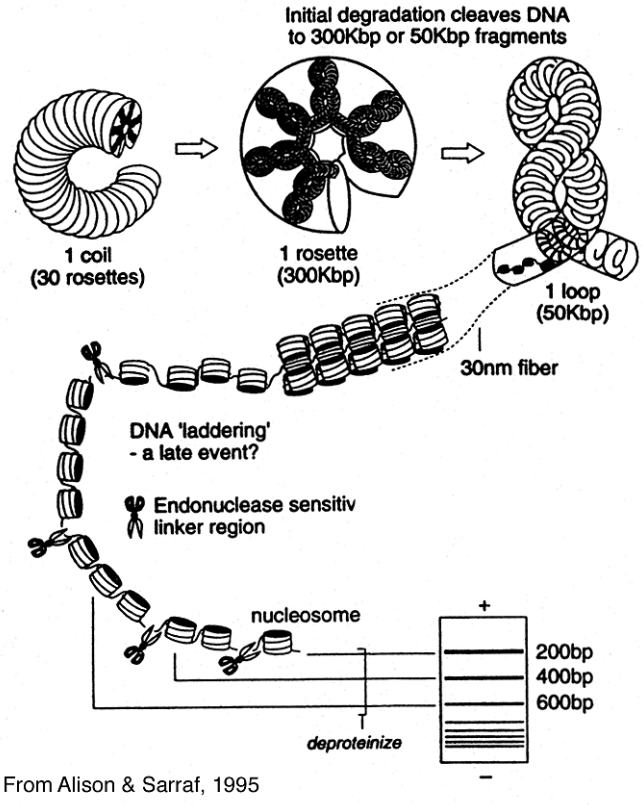


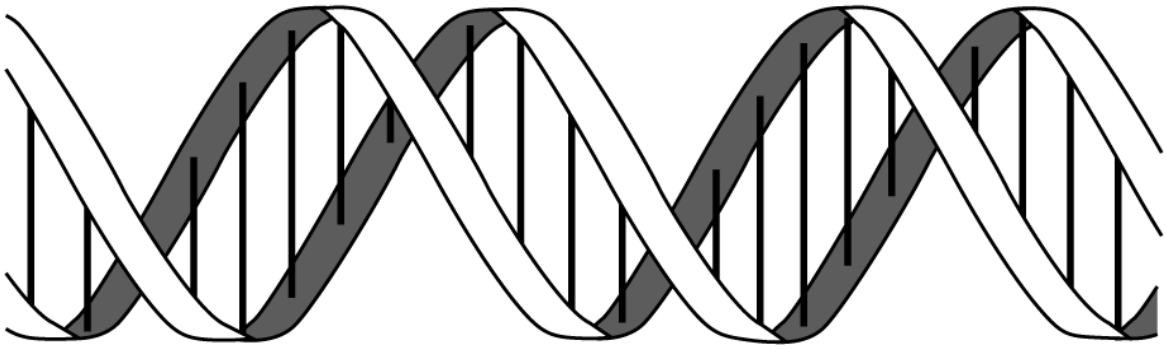
Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Internucleosomal cleavage of DNA occurs during the process of apoptosis and electrophoresis of the resulting DNA fragments produces the DNA ladder pattern that is a hallmark of apoptosis.

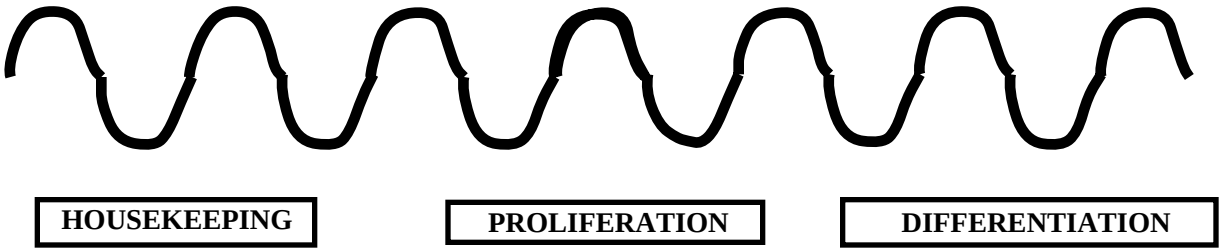


From Alison & Sarraf, 1995

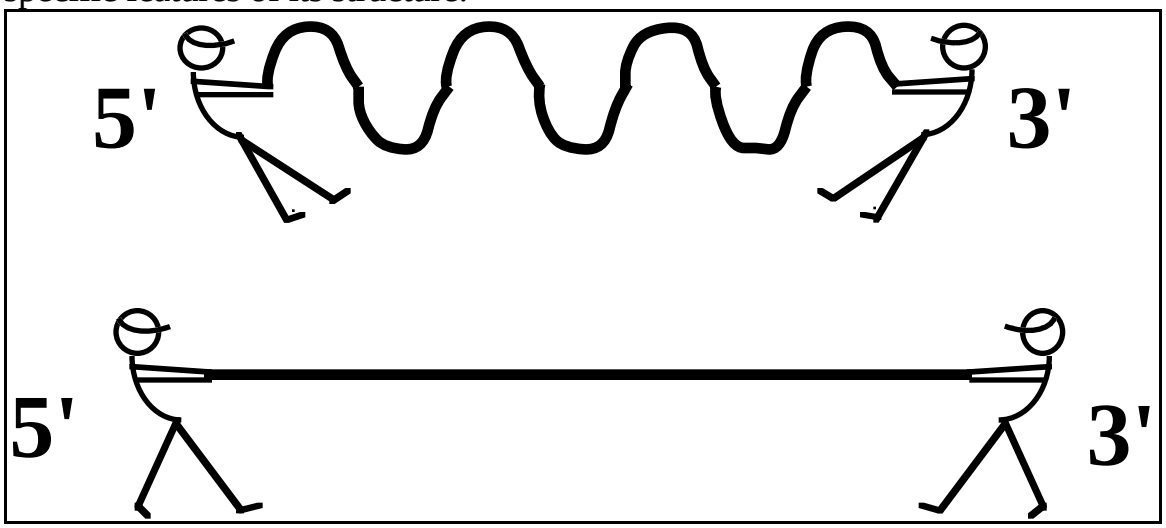
DNA exists as a double helix with the two parts joined by hydrogen bonds.

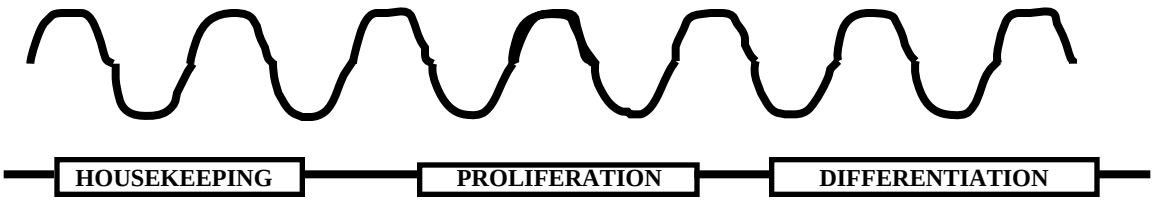


Different genes are arranged along the length of each strand of the double helix.



In conventional diagrams, DNA strands are straightened out to help illustrate specific features of its structure.





A gene includes regions that precede and follow a more centrally located **coding** region. Genes for polypeptides include a **leader** region followed by the coding region followed by the **trailer**.

Leader -----> coding region -----> trailer

Intervening sequences separate genes on a strand of DNA.

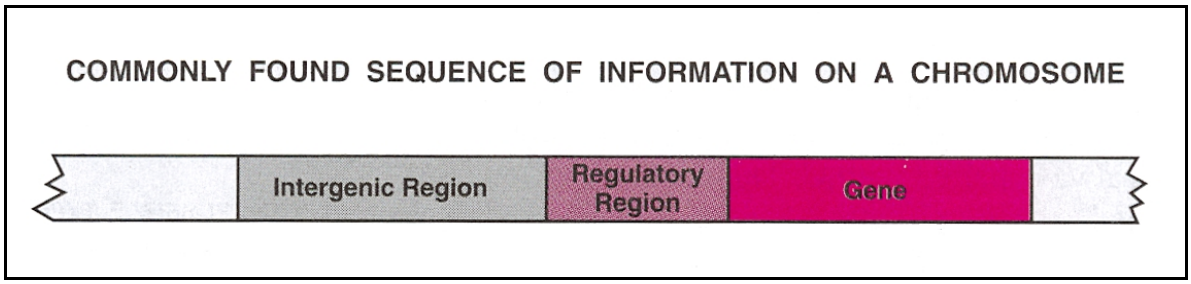


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

The leader and trailer are not translated into protein. The coding region is divided into exons and introns.

Leader -----> coding region -----> trailer

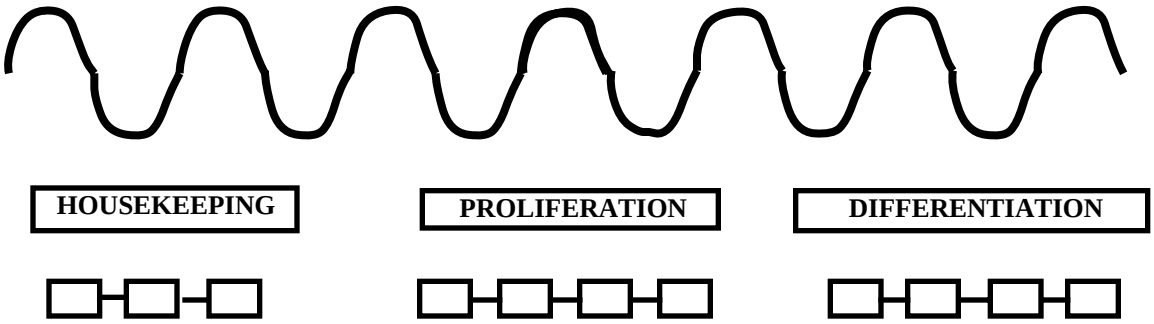
exons --- introns

Exons encode or **ex**press specific amino acids that ultimately make up the polypeptide gene product. **Introns int**errupt or **int**ervene between exons. While introns are transcribed into immature RNA, they are removed from

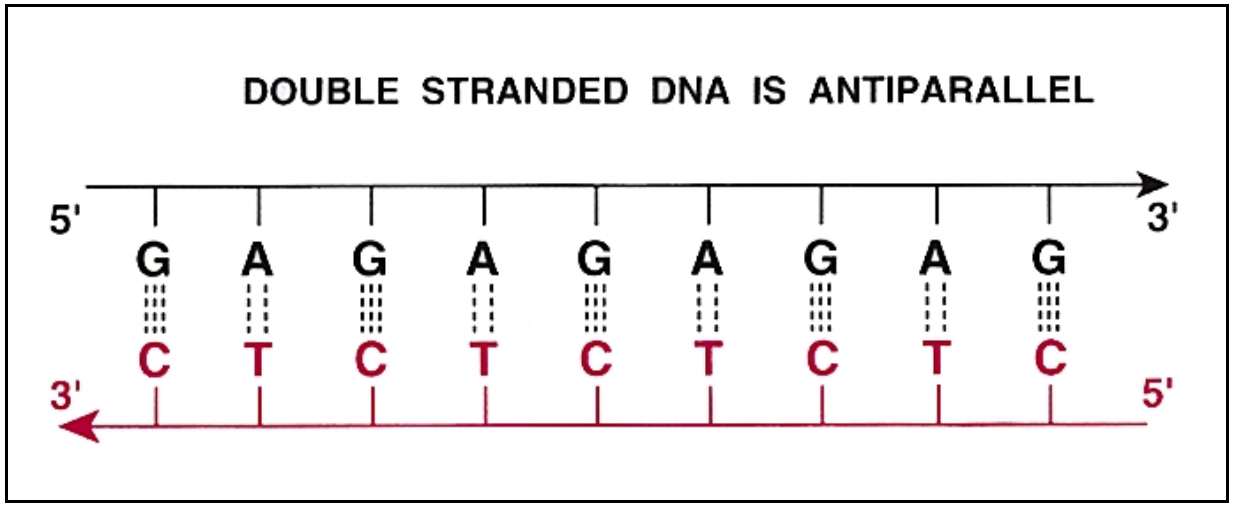
within the transcript by **splicing** together of exons on either side and, thus, do not encode amino acids.

Leader ----(intron---exon---intron---exon---intron)$_n$----trailer

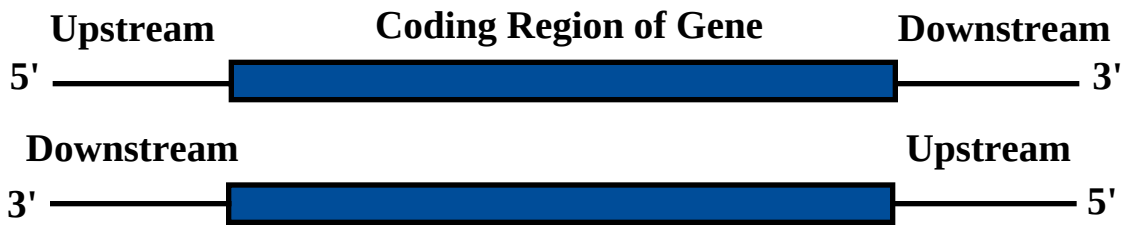
The separate strands of DNA are defined as a coding strand and a template strand.



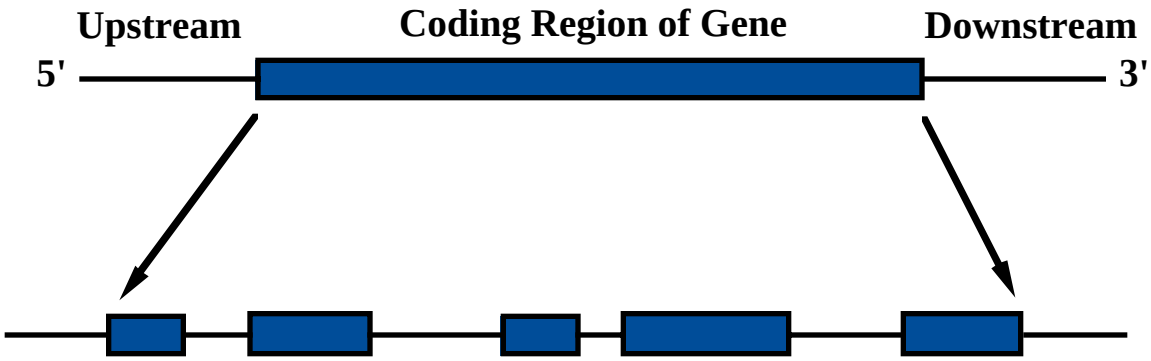
The genes have polarity characterized by a 5' upstream end and a 3' downstream end. In its double stranded form the complementary strands are antiparallel.



The following 3 illustrations are meant to summarize what has been presented relative to gene structure.



Template Region of Gene



= Exons **= Introns**

Exon = Expressed portion of the DNA
Intron = Intervening portion of the DNA

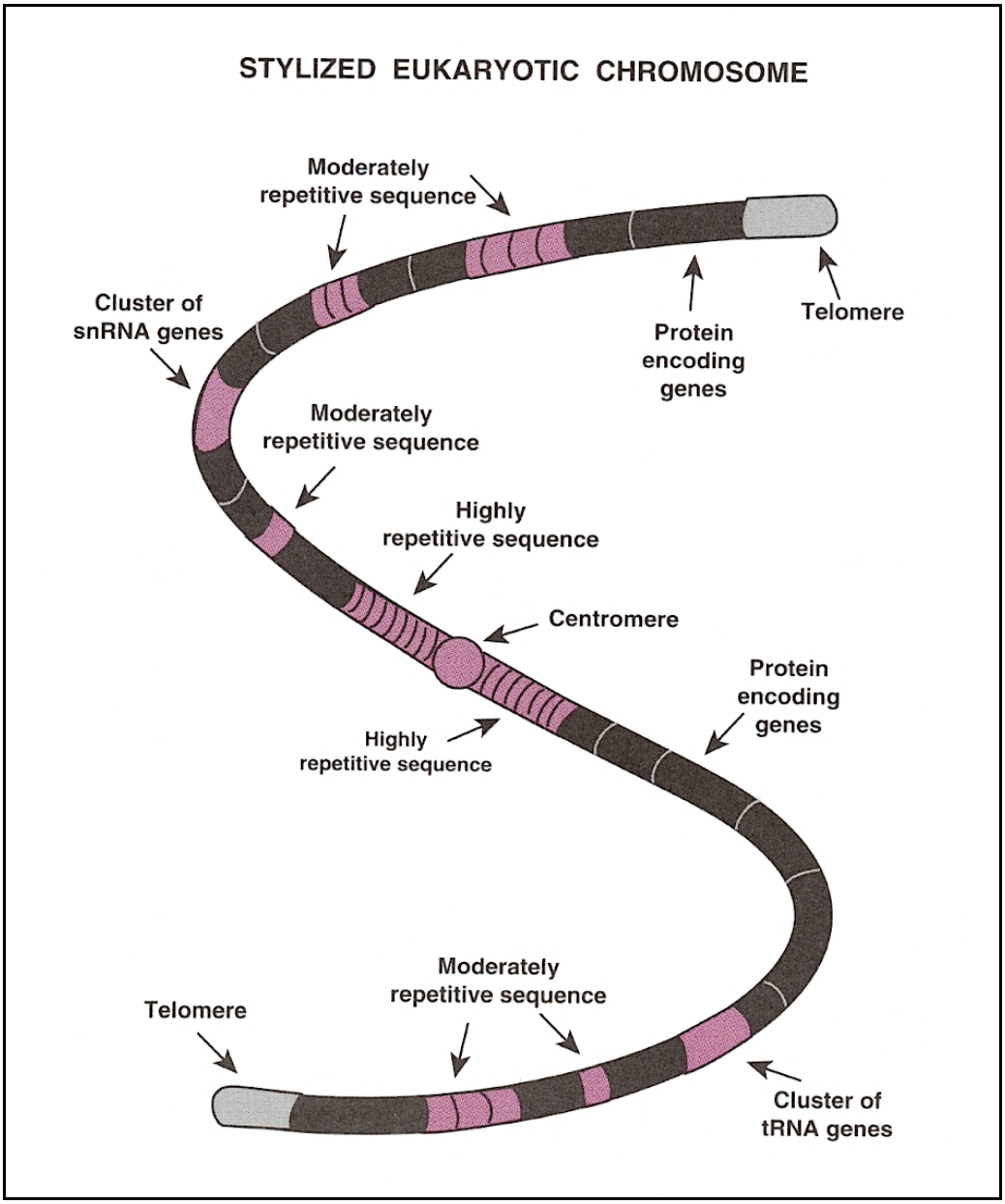


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, IIllinois, USA.

Repetitive sequences - Hundreds to thousands of copies of repetitive sequences occur both within genes, possibly representing multiple copies of highly used genes, and between genes, possibly serving a spacers. Their exact function is unknown. One commonly occurring highly repetitive sequence is the *Alu* sequence (*Alu* element) consisting of 300,000 to 500,000 base pairs scattered throughout the human genome. These contain a single site for the restriction endonuclease *Alu*I.

snRNA = small nuclear RNA; involved in splicing of mRNA.

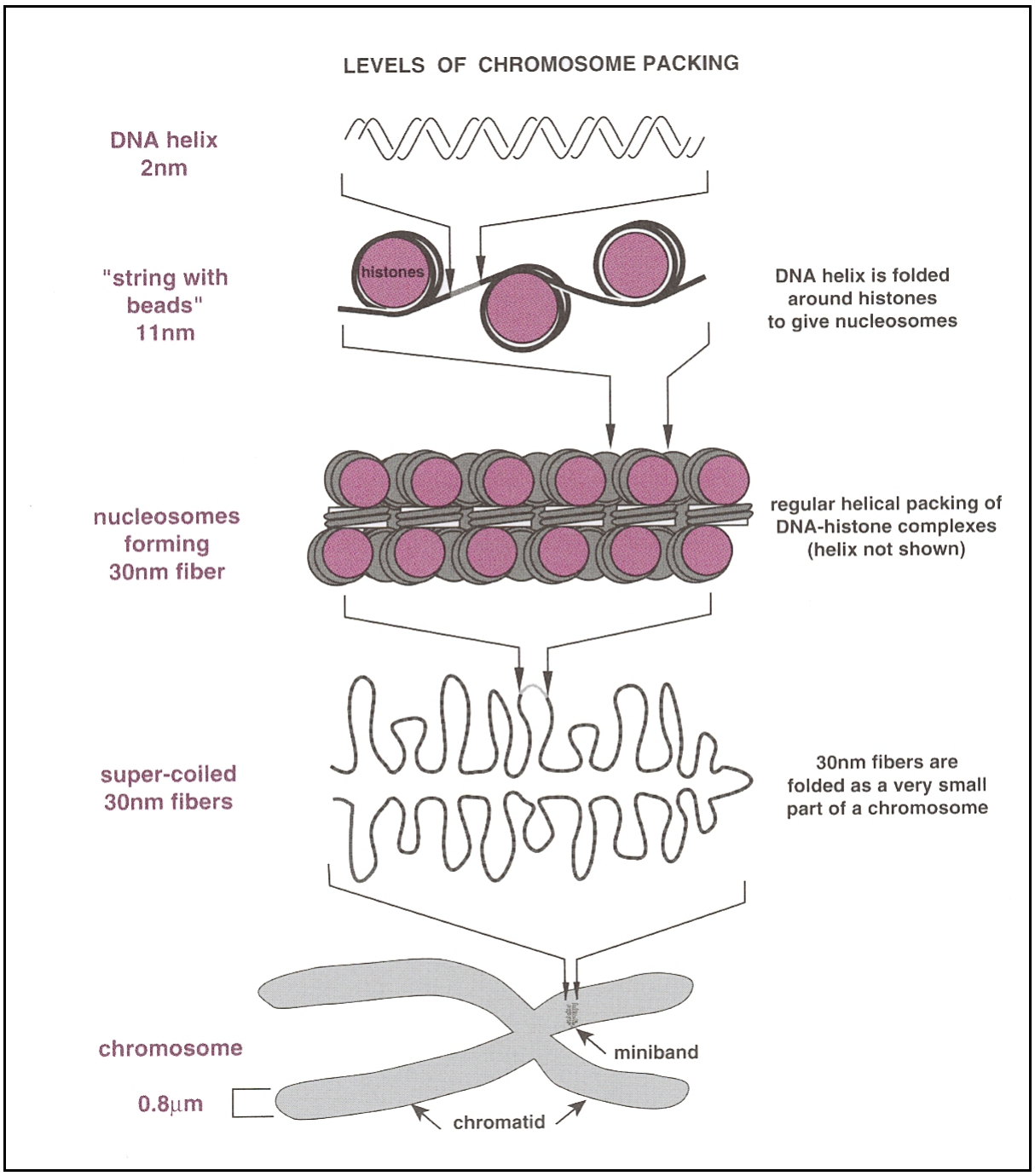


Figure from: D.P. Clark and L.D. Russell (1997) *Molecular Biology Made Fun and Simple* with permission from Cache River Press, Vienna, Illlinois, USA.